<Bioinformatics Center> Biological Information Networks

Bioinformatics Center
- Biological Information Networks -

http://www.bic.kyoto-u.ac.jp/takutsu/index.html

Prof
AKUTSU, Tatsuya
(D Eng)

Assist Prof
HAYASHIDA, Morihiro
(D Inf)

Assist Prof
TAMURA, Takeyuki
(D Inf)

PD (JSPS)
KATO, Yuki
(D Eng)

Guest Res Assoc
SONG, Jiangning
(Ph D)

Students

MOURI, Kazunari (D2)
TAKEMOTO, Kazuhiro (D2)
BROWN, John (D1)
POOLSAP, Unyanee (M2)
CHO, Uncho (M2)
URATA, Takashi (M2)

Visitors

Assoc Prof DAI, Yang — University of Illinois at Chicago, USA, 23 April 2007
TAN, Hao — Monash University, Australia, 7 December 2007–1 March 2008

Scope of Research

Due to rapid progress of the genome projects, whole genome sequences of organisms ranging from bacteria to human have become available. In order to understand the meaning behind the genetic code, we have been developing algorithms and software tools for analyzing biological data based on advanced information technologies such as theory of algorithms, artificial intelligence, and machine learning. We are recently studying the following topics: systems biology, scale-free networks, protein structure prediction, inference of biological networks, chemo-informatics, discrete and stochastic methods for bioinformatics.

Research Activities (Year 2007)

Publications

Takemoto K, Nacher JC, Akutsu T: Correlation between Structure and Temperature in Prokaryotic Metabolic Networks, *BMC Bioinformatics*, **8**, [303-1]-[303-11] (2007).

Akutsu T, Hayashida M, Ching WK, Ng MK: Control of Boolean Networks: Hardness Results and Algorithms for Tree Structured Networks, *Journal of Theoretical Biology*, **244**, 670-679 (2007).

Tamura T, Akutsu T: Subcellular Location Prediction of Proteins Using Support Vector Machines with Alignment of Block Sequences Utilizing Amino Acid Composition, *BMC Bioinformatics*, **8**, 466 (2007).

Presentations

An $O(1.787^n)$-time Algorithm for Detecting a Singleton Attractor in a Boolean Network Consisting of AND/OR Nodes, Tamura T, Akutsu T, The 16th International Symposium on Fundamentals of Computation Theory, 30 August 2007.

A Grammatical Approach to RNA-RNA Interaction Prediction, Kato Y, Akutsu T, Seki H (Nara Institute of Science and Technology), International Symposium on Computational Models for Life Sciences, 17 December 2007.

Grants

Akutsu T, Kawabata T, Nagamochi H, Hayashida M, A Novel Approach to Computational Drug Design Based on Graph Theory and Kernel Methods, Grant-in-Aid for Scientific Research (A), 1 April 2007–31 March 2010.

Akutsu T, Goto S, Mochizuki A, Tokita K, Mathematical Analysis of Structure and Dynamics of Biological Information Networks, Grant-in-Aid for Scientific Research on Priority Areas, 1 April 2005–31 March 2010.

Awards

Akutsu T, Contribution Award, Special Interest Group on Mathematical Modeling and Problem Solving, Information Processing Society of Japan, 3 March 2007.

Tamura T, FIT Award for Young Researchers, Approximation Algorithms for Optimal RNA Secondary Structures Common to Multiple Sequences, IPSJ and IEICE, 6 September 2007.

Structural Difference with Temperature in Prokaryotic Metabolic Networks

Organisms grow in the environment of different temperatures. Since heat-loving organisms are believed to be primeval forms of life, elucidation of differences with temperature is a major topic in evolutionary biology. Up until now, the adaptive differences as a result of temperature have been revealed in structural and sequence properties of transcriptomes and proteomes. However, temperature-dependent differences in metabolism are still unclear.

We here represent the metabolism as a network (graph) in which nodes and edges correspond to metabolites and substrate-product relationships between them, respectively, and investigate a relationship between structure and optimal growth temperature in metabolic networks of 113 prokaryotes using graph-theoretical metrics. As a result, we find significant correlations between structural properties and optimal growth temperature in the metabolic networks (e.g. see Figure 1A). The metabolic networks become less dense and low modular with increasing temperature. Furthermore, the connectivity of the networks becomes homogenous. This result implies that metabolic networks undergo a change from ordered structures such as clustered scale-free networks to disordered structures such as random networks with increasing temperature (Figure 1B). Our finding might suggest that the temperature plays an important role in design principles of metabolic networks.

Takemoto K, Nacher J C, Akutsu T, *BMC Bioinformatics*, **8**, 303 (2007).

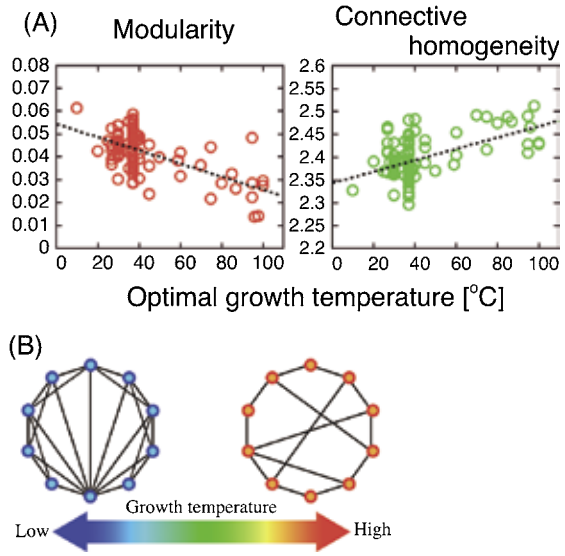

Figure 1. (A) Major example of statistically significant correlations between structure properties and optimal growth temperature. (B) Schematic diagram of a structural transition of metabolic networks with temperature.

Subcellular Location Prediction of Proteins Using Support Vector Machines with Alignment of Block Sequences Utilizing Amino Acid Composition

Subcellular location prediction of proteins is an important and well-studied problem in bioinformatics.

This is a problem of predicting which part in a cell a given protein is transported to, where an amino acid sequence of the protein is given as an input. This problem is becoming more important since information on subcellular location is helpful for annotation of proteins and genes and the number of complete genomes is rapidly increasing. Since existing predictors are based on various heuristics, it is important to develop a simple method with high prediction accuracies.

In this work, we propose a novel and general predicting method by combining techniques for sequence alignment and feature vectors based on amino acid composition. We implemented this method with support vector machines on plant data sets extracted from the TargetP database.

Through fivefold cross validation tests, the obtained overall accuracies and average MCC were 0.9096 and 0.8655 respectively. We also applied our method to other datasets including that of WoLF PSORT.

Although there is a predictor which uses the information of gene ontology and yields higher accuracy than ours, our accuracies are higher than existing predictors which use only sequence information. Since such information as gene ontology can be obtained only for known proteins, our predictor is considered to be useful for subcellular location prediction of newly-discovered proteins. Furthermore, the idea of combination of alignment and amino acid frequency is novel and general so that it may be applied to other problems in bioinformatics. Our method for plant is also implemented as a web-system and available on [http://sunflower.kuicr.kyoto-u.ac.jp/~tamura/slpfa.html].

Tamura T, Akutsu T: *BMC Bioinformatics*, **8**, 466 (2007)

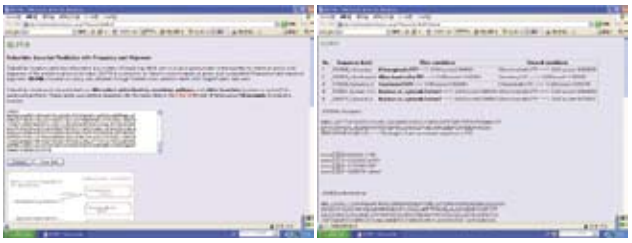

Figure 2. (Left) SLPFA prediction server. (Right) Result of an example.